

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение высшего образования
"Казанский (Приволжский) федеральный университет"
Институт вычислительной математики и информационных технологий



подписано электронно-цифровой подписью

Программа дисциплины

Анализ данных на Python

Направление подготовки: 01.04.02 - Прикладная математика и информатика

Профиль подготовки: Искусственный интеллект и суперкомпьютерные вычисления

Квалификация выпускника: магистр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2024

Содержание

1. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения ОПОП ВО
2. Место дисциплины (модуля) в структуре ОПОП ВО
3. Объем дисциплины (модуля) в зачетных единицах с указанием количества часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся
4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий
 - 4.1. Структура и тематический план контактной и самостоятельной работы по дисциплине (модулю)
 - 4.2. Содержание дисциплины (модуля)
5. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)
6. Фонд оценочных средств по дисциплине (модулю)
7. Перечень литературы, необходимой для освоения дисциплины (модуля)
8. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)
9. Методические указания для обучающихся по освоению дисциплины (модуля)
10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем (при необходимости)
11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)
12. Средства адаптации преподавания дисциплины (модуля) к потребностям обучающихся инвалидов и лиц с ограниченными возможностями здоровья
13. Приложение №1. Фонд оценочных средств
14. Приложение №2. Перечень литературы, необходимой для освоения дисциплины (модуля)
15. Приложение №3. Перечень информационных технологий, используемых для освоения дисциплины (модуля), включая перечень программного обеспечения и информационных справочных систем

Программу дисциплины разработал(а)(и): старший преподаватель, б/с Гиниятова Д.Х. (кафедра прикладной математики и искусственного интеллекта, отделение прикладной математики и информатики), dginiyat@kpfu.ru

1. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения ОПОП ВО

Обучающийся, освоивший дисциплину (модуль), должен обладать следующими компетенциями:

| Шифр компетенции | Расшифровка приобретаемой компетенции |
|------------------|--|
| ПК-4 | Разработка, отладка, рефакторинг программного кода, баз данных, информационных ресурсов; проектирование и интеграция программного обеспечения, управление проектами в области ИТ |
| ПК-5 | Разработка, совершенствование и внедрение новых методов, технологий и инструментальных средств для получения, обработки, использования и аналитики больших данных |

Обучающийся, освоивший дисциплину (модуль):

Должен знать:

- основы языка Python
- теоретические аспекты анализа данных
- специализированные библиотеки для визуализации, предварительной обработки и анализа данных на Python

Должен уметь:

- осуществлять поиск, агрегацию данных по теме исследования
- проводить предварительный анализ инструментами Python
- интерпретировать полученные результаты и делать на их основе выводы

Должен владеть:

- владеть инструментами визуализации, обработки и анализа данных средствами Python

2. Место дисциплины (модуля) в структуре ОПОП ВО

Данная дисциплина (модуль) включена в раздел "Б1.В.ДВ.02.01 Дисциплины (модули)" основной профессиональной образовательной программы 01.04.02 "Прикладная математика и информатика (Искусственный интеллект и суперкомпьютерные вычисления)" и относится к дисциплинам по выбору части ОПОП ВО, формируемой участниками образовательных отношений.

Осваивается на 2 курсе в 3 семестре.

3. Объем дисциплины (модуля) в зачетных единицах с указанием количества часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся

Общая трудоемкость дисциплины составляет 3 зачетных(ые) единиц(ы) на 108 часа(ов).

Контактная работа - 18 часа(ов), в том числе лекции - 0 часа(ов), практические занятия - 0 часа(ов), лабораторные работы - 18 часа(ов), контроль самостоятельной работы - 0 часа(ов).

Самостоятельная работа - 90 часа(ов).

Контроль (зачёт / экзамен) - 0 часа(ов).

Форма промежуточного контроля дисциплины: зачет в 3 семестре.

4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

4.1 Структура и тематический план контактной и самостоятельной работы по дисциплине (модулю)

| N | Разделы дисциплины / модуля | Се- местр | Виды и часы контактной работы, их трудоемкость (в часах) | | | | | | Само- стоя- тель- ная ра- бота |
|----|---|--------------|--|--------------------------|--|-------------------------------------|---------------------------------------|-------------------------------------|---|
| | | | Лекции, всего | Лекции в эл. форме | Практи- ческие занятия, всего | Практи- ческие в эл. форме | Лабора- торные работы, всего | Лабора- торные в эл. форме | |
| N | Разделы дисциплины / модуля | Се- местр | Виды и часы контактной работы, их трудоемкость (в часах) | | | | | | Само- стоя- тель- ная ра- бота |
| | | | Лекции, всего | Лекции в эл. форме | Практи- ческие занятия, всего | Практи- ческие в эл. форме | Лабора- торные работы, всего | Лабора- торные в эл. форме | |
| 1. | Тема 1. Основные задачи анализа данных. Основы работы с данными в Python | 3 | 0 | 0 | 0 | 0 | 4 | 0 | 30 |
| 2. | Тема 2. Обработка данных в Python | 3 | 0 | 0 | 0 | 0 | 8 | 0 | 30 |
| 3. | Тема 3. Визуализация данных в Python. Работа с API. | 3 | 0 | 0 | 0 | 0 | 6 | 0 | 30 |
| 4. | Тема 4. Основные задачи анализа данных. Основы работы с данными в Python. Введение в анализ данных на Python. Основы синтаксиса Python. Основные алгоритмические конструкции. Структуры данных. Работа со строковыми данными, регулярными выражениями и текстовыми файлами. Работа с файлами и сложными метриками. Введение в Pandas - обзорное занятие по работе с библиотекой Pandas. | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 90 |

Тема 2. Обработка данных в Python

Обработка данных в Pandas. Простые расчеты в таблицах. Строковые и lambda-функции. Сложные функции для датафреймов и работы с ссылками. Группировки данных. Агрегированные группировки датафреймов. Сводные таблицы в Pandas. Агрегированные столбцы датафреймов. Дополнительные методы работы с данными: очистка данных и создание новых признаков ("feature engineering"). Объединение датафреймов. Работа с множеством файлов.

Тема 3. Визуализация данных в Python. Работа с API.

Работа с датами. Распознавание дат. Прибавление интервала и строки. Перебор дат в цикле. Расчет lifetime. Графические возможности Pandas. Оформление графиков matplotlib. Seaborn для продвинутых визуализаций. Plotly для интерактивных отчетов. Парсинг сайтов и основы HTML. Работа с API на примере API Вконтакте. API сервисов Яндекса. Работа с ошибками и ботами. Обработка естественного языка и регулярные выражения.

5. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)

Самостоятельная работа обучающихся выполняется по заданию и при методическом руководстве преподавателя, но без его непосредственного участия. Самостоятельная работа подразделяется на самостоятельную работу на аудиторных занятиях и на внеаудиторную самостоятельную работу. Самостоятельная работа обучающихся включает как полностью самостоятельное освоение отдельных тем (разделов) дисциплины, так и проработку тем (разделов), осваиваемых во время аудиторной работы. Во время самостоятельной работы обучающиеся читают и конспектируют учебную, научную и справочную литературу, выполняют задания, направленные на закрепление знаний и отработку умений и навыков, готовятся к текущему и промежуточному контролю по дисциплине.

Организация самостоятельной работы обучающихся регламентируется нормативными документами, учебно-методической литературой и электронными образовательными ресурсами, включая:

Порядок организации и осуществления образовательной деятельности по образовательным программам высшего образования - программам бакалавриата, программам специалитета, программам магистратуры (утвержден приказом Министерства науки и высшего образования Российской Федерации от 6 апреля 2021 года №245)

Письмо Министерства образования Российской Федерации №14-55-99бин/15 от 27 ноября 2002 г. "Об активизации самостоятельной работы студентов высших учебных заведений"

Устав федерального государственного автономного образовательного учреждения "Казанский (Приволжский) федеральный университет"

Правила внутреннего распорядка федерального государственного автономного образовательного учреждения высшего профессионального образования "Казанский (Приволжский) федеральный университет"

Локальные нормативные акты Казанского (Приволжского) федерального университета

6. Фонд оценочных средств по дисциплине (модулю)

Фонд оценочных средств по дисциплине (модулю) включает оценочные материалы, направленные на проверку освоения компетенций, в том числе знаний, умений и навыков. Фонд оценочных средств включает оценочные средства текущего контроля и оценочные средства промежуточной аттестации.

В фонде оценочных средств содержится следующая информация:

- соответствие компетенций планируемым результатам обучения по дисциплине (модулю);
- критерии оценивания сформированности компетенций;
- механизм формирования оценки по дисциплине (модулю);
- описание порядка применения и процедуры оценивания для каждого оценочного средства;
- критерии оценивания для каждого оценочного средства;
- содержание оценочных средств, включая требования, предъявляемые к действиям обучающихся, демонстрируемым результатам, задания различных типов.

Фонд оценочных средств по дисциплине находится в Приложении 1 к программе дисциплины (модулю).

7. Перечень литературы, необходимой для освоения дисциплины (модуля)

Освоение дисциплины (модуля) предполагает изучение основной и дополнительной учебной литературы. Литература может быть доступна обучающимся в одном из двух вариантов (либо в обоих из них):

- в электронном виде - через электронные библиотечные системы на основании заключенных КФУ договоров с правообладателями;

- в печатном виде - в Научной библиотеке им. Н.И. Лобачевского. Обучающиеся получают учебную литературу на абонементе по читательским билетам в соответствии с правилами пользования Научной библиотекой.

Электронные издания доступны дистанционно из любой точки при введении обучающимся своего логина и пароля от личного кабинета в системе "Электронный университет". При использовании печатных изданий библиотечный фонд должен быть укомплектован ими из расчета не менее 0,5 экземпляра (для обучающихся по ФГОС 3++ - не менее 0,25 экземпляра) каждого из изданий основной литературы и не менее 0,25 экземпляра дополнительной литературы на каждого обучающегося из числа лиц, одновременно осваивающих данную дисциплину.

Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля), находится в Приложении 2 к рабочей программе дисциплины. Он подлежит обновлению при изменении условий договоров КФУ с правообладателями электронных изданий и при изменении комплектования фондов Научной библиотеки КФУ.

8. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)

Изучение и анализа данных с помощью Python -

<https://learn.microsoft.com/ru-ru/training/modules/explore-analyze-data-with-python/>

Официальная документация Pandas - <https://pandas.pydata.org/docs/>

Официальная документация Python - <https://docs.python.org/3/>

9. Методические указания для обучающихся по освоению дисциплины (модуля)

| Вид работ | Методические рекомендации |
|------------------------|---|
| лабораторные работы | Для успешного прохождения курса необходимо предварительно изучить теоретический материал по каждой теме, а затем проделать все операции, команды, вызвать функции, указанные в теоретическом материале, рекомендуется обращаться к официальной документации по библиотекам Python. Выполнять практические задания необходимо по каждой теме, так как каждый следующий материал, основан на предыдущем |
| самостоятельная работа | Самостоятельная работа является обязательной частью образовательного процесса и включает в себя самостоятельную проработку теоретического материала, чтение документации по необходимым библиотекам Python, более детальное рассмотрение вопросов, рассмотренных на практическом занятии. Самостоятельная работа должна проводиться систематически. |
| зачет | Для успешной сдачи зачета необходима систематическая подготовка и проработка материалов практических занятий. Перед выполнением задания, вынесенного на зачет, необходимо внимательно ознакомиться с текстом задания, с данными, предоставленными для его выполнения, и что конкретно требуется определить после анализа данных. |

10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем (при необходимости)

Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем, представлен в Приложении 3 к рабочей программе дисциплины (модуля).

11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

Материально-техническое обеспечение образовательного процесса по дисциплине (модулю) включает в себя следующие компоненты:

Помещения для самостоятельной работы обучающихся, укомплектованные специализированной мебелью (столы и стулья) и оснащенные компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду КФУ.

Учебные аудитории для контактной работы с преподавателем, укомплектованные специализированной мебелью (столы и стулья).

Компьютер и принтер для распечатки раздаточных материалов.

12. Средства адаптации преподавания дисциплины к потребностям обучающихся инвалидов и лиц с ограниченными возможностями здоровья

При необходимости в образовательном процессе применяются следующие методы и технологии, облегчающие восприятие информации обучающимися инвалидами и лицами с ограниченными возможностями здоровья:

- создание текстовой версии любого нетекстового контента для его возможного преобразования в альтернативные формы, удобные для различных пользователей;
- создание контента, который можно представить в различных видах без потери данных или структуры, предусмотреть возможность масштабирования текста и изображений без потери качества, предусмотреть доступность управления контентом с клавиатуры;
- создание возможностей для обучающихся воспринимать одну и ту же информацию из разных источников - например, так, чтобы лица с нарушениями слуха получали информацию визуально, с нарушениями зрения - аудиально;
- применение программных средств, обеспечивающих возможность освоения навыков и умений, формируемых дисциплиной, за счёт альтернативных способов, в том числе виртуальных лабораторий и симуляционных технологий;
- применение дистанционных образовательных технологий для передачи информации, организации различных форм интерактивной контактной работы обучающегося с преподавателем, в том числе вебинаров, которые могут быть использованы для проведения виртуальных лекций с возможностью взаимодействия всех участников дистанционного обучения, проведения семинаров, выступления с докладами и защиты выполненных работ, проведения тренингов, организации коллективной работы;
- применение дистанционных образовательных технологий для организации форм текущего и промежуточного контроля;
- увеличение продолжительности сдачи обучающимся инвалидом или лицом с ограниченными возможностями здоровья форм промежуточной аттестации по отношению к установленной продолжительности их сдачи:
- продолжительности сдачи зачёта или экзамена, проводимого в письменной форме, - не более чем на 90 минут;
- продолжительности подготовки обучающегося к ответу на зачёте или экзамене, проводимом в устной форме, - не более чем на 20 минут;
- продолжительности выступления обучающегося при защите курсовой работы - не более чем на 15 минут.

Программа составлена в соответствии с требованиями ФГОС ВО и учебным планом по направлению 01.04.02 "Прикладная математика и информатика" и магистерской программе "Искусственный интеллект и суперкомпьютерные вычисления".

Приложение 2
к рабочей программе дисциплины (модуля)
Б1.В.ДВ.02.01 Анализ данных на Python

Перечень литературы, необходимой для освоения дисциплины (модуля)

Направление подготовки: 01.04.02 - Прикладная математика и информатика
Профиль подготовки: Искусственный интеллект и суперкомпьютерные вычисления
Квалификация выпускника: магистр
Форма обучения: очное
Язык обучения: русский
Год начала обучения по образовательной программе: 2024

Основная литература:

1. Баяк, О. А. Практикум по анализу данных на языках Python и R : учебное пособие / О. А. Баяк, М. Р. Исаева, М. О. Самсонкин. - Москва: Прометей, 2023. - 100 с. - ISBN 978-5-00172-356-1. - Текст : электронный. - URL: <https://znanium.com/catalog/product/2124862> (дата обращения: 12.01.2024). - Режим доступа: по подписке.
2. Маккинли, У. Python и анализ данных / Уэс Маккинли - Москва : ДМК Пресс, 2015. - 482 с. - ISBN 978-5-97060-315-4. - Текст : электронный // ЭБС 'Консультант студента' : [сайт]. - URL : <https://www.studentlibrary.ru/book/ISBN9785970603154.html> (дата обращения: 12.01.2024). - Режим доступа : по подписке.
3. Малов Д. А. Глубокое обучение и анализ данных. Практическое руководство / Д.А. Малов. - Санкт-Петербург : БХВ-Петербург, 2023. - 272 с. - ISBN 978-5-9775-1172-8. - URL: <https://ibooks.ru/bookshelf/389629> (дата обращения: 12.01.2024). - Текст: электронный.

Дополнительная литература:

1. Хейдт, М. Изучаем pandas / Хейдт М. , пер. с англ. А. В. Груздева. - Москва : ДМК Пресс, 2018. - 438 с. - ISBN 978-5-97060-625-4. - Текст : электронный // ЭБС 'Консультант студента' : [сайт]. - URL : <https://www.studentlibrary.ru/book/ISBN9785970606254.html> (дата обращения: 12.01.2024). - Режим доступа : по подписке.
2. Tandon Akash. Расширенная аналитика с PySpark: Пер. с англ. - Санкт-Петербург : БХВ-Петербург, 2023. - 224 с. - ISBN 978-5-9775-1770-6. - URL: <https://ibooks.ru/bookshelf/389656> (дата обращения: 12.01.2024). - Текст: электронный.
3. Нисчал Н. Python - это просто. Пошаговое руководство по программированию и анализу данных: Пер. с англ. / Н. Нисчал. - Санкт-Петербург : БХВ-Петербург, 2022. - 416 с. - ISBN 978-5-9775-6849-4. - URL: <https://ibooks.ru/bookshelf/386516> (дата обращения: 12.01.2024). - Текст: электронный.

Перечень информационных технологий, используемых для освоения дисциплины (модуля), включая перечень программного обеспечения и информационных справочных систем

Направление подготовки: 01.04.02 - Прикладная математика и информатика

Профиль подготовки: Искусственный интеллект и суперкомпьютерные вычисления

Квалификация выпускника: магистр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2024

Освоение дисциплины (модуля) предполагает использование следующего программного обеспечения и информационно-справочных систем:

Операционная система Microsoft Windows 7 Профессиональная или Windows XP (Volume License)

Пакет офисного программного обеспечения Microsoft Office 365 или Microsoft Office Professional plus 2010

Браузер Mozilla Firefox

Браузер Google Chrome

Adobe Reader XI или Adobe Acrobat Reader DC

Kaspersky Endpoint Security для Windows

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе "ZNANIUM.COM", доступ к которой предоставлен обучающимся. ЭБС "ZNANIUM.COM" содержит произведения крупнейших российских учёных, руководителей государственных органов, преподавателей ведущих вузов страны, высококвалифицированных специалистов в различных сферах бизнеса. Фонд библиотеки сформирован с учетом всех изменений образовательных стандартов и включает учебники, учебные пособия, учебно-методические комплексы, монографии, авторефераты, диссертации, энциклопедии, словари и справочники, законодательно-нормативные документы, специальные периодические издания и издания, выпускаемые издательствами вузов. В настоящее время ЭБС ZNANIUM.COM соответствует всем требованиям федеральных государственных образовательных стандартов высшего образования (ФГОС ВО) нового поколения.

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе Издательства "Лань", доступ к которой предоставлен обучающимся. ЭБС Издательства "Лань" включает в себя электронные версии книг издательства "Лань" и других ведущих издательств учебной литературы, а также электронные версии периодических изданий по естественным, техническим и гуманитарным наукам. ЭБС Издательства "Лань" обеспечивает доступ к научной, учебной литературе и научным периодическим изданиям по максимальному количеству профильных направлений с соблюдением всех авторских и смежных прав.

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе "Консультант студента", доступ к которой предоставлен обучающимся. Многопрофильный образовательный ресурс "Консультант студента" является электронной библиотечной системой (ЭБС), предоставляющей доступ через сеть Интернет к учебной литературе и дополнительным материалам, приобретенным на основании прямых договоров с правообладателями. Полностью соответствует требованиям федеральных государственных образовательных стандартов высшего образования к комплектованию библиотек, в том числе электронных, в части формирования фондов основной и дополнительной литературы.